DOCUMENT RESUME

ED 249 780 FL 014 606

AUTHOR Morgan, James L.
TITLE Some Concepts and Consequences of a Degree 1 Learnability Result.
INSTITUTION Stanford Univ., Calif. Dept. of Linguistics.
PUB DATE Sep 84
NOTE 9p.; In: Papers and Reports on Child Language Development, Volume 23, p98-105 Sep 1984.
AVAILABLE FROM PRCLD, Department of Linguistics, Stanford University, Stanford, CA 94305 ($12.00 for entire volume; individual papers not available).
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Child Language; *Difficulty Level; *Grammar; *Language Acquisition; Learning Processes; *Learning Theories; Mathematical Formulas; Models
IDENTIFIERS *Learnability Theory

ABSTRACT

Learnability theory involves the construction of formal mathematical proofs whose goal is to demonstrate how the child can successfully induce a mature grammar. An empirically adequate learnability proof constitutes a detailed hypothesis concerning the boundary conditions within which acquisition proceeds and can provide a general framework for developmental accounts of specific aspects of language acquisition. A central difficulty in constructing a feasible learnability proof for natural languages is that the child's language input is strictly limited. One line of inquiry has been concerned with demonstrating how transformational grammars might in principle be learned, given strict constraints on the complexity of available input. Degree 1 learnability involves learning from constructions with one level of embedding, less complex constructions than have been used for Degree 2 learnability proofs. The Degree 1 proof avoids some of the serious problems encountered in Degree 2 proofs, and in the result outlined it is shown that in the acquisition of complex grammars such as those of natural languages, the required complexity of input can be reduced and constraints on possible grammatical hypotheses can simultaneously be relaxed if the input available to the language learner is rich in structural information. (Author/MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

PRCLD #23
:984

ED249780

Some Concepts and Consequences of a Degree 1 Learnability Result

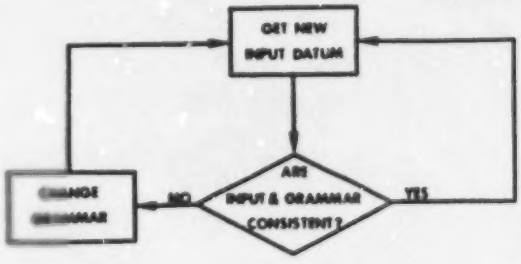
James L. Morgan
University of Minnesota

Learnability theory involves the construction of formal, mathematical proofs whose goal is to demonstrate how the child can in principle successfully induce a mature grammar. Once a proof is constructed, the assumptions incorporated in it are typically tested for their consistency with natural language phenomena. An empirically adequate learnability proof constitutes a detailed hypothesis concerning the boundary conditions within which acquisition proceeds and thus can provide a general framework for developmental accounts of specific aspects of language acquisition.

As a general model of how acquisition progresses, most work in learnability has adopted a system similar to that shown in Figure 1. In each cycle, the child, or learner, receives a bit of input data and determines whether this new datum and her internal grammar are consistent. If so, nothing need be done until the next input datum is received. However, if input and grammar are not consistent, the learner must hypothesize some modification to her grammar. Then the cycle begins again. The learner is said to have succeeded if, first, after some point in time, the learner never has to change her grammar again, and, second, if the learner's grammar then generates the language from which her input has been drawn. These requirements constitute a criterion of learning (Gold, 1967).

Figure 1.

A Learning System.



In addition to this criterion, there are three components for which a learnability proof must supply explicit definitions. First, input must be characterized. Minimally, input must include some number of utterances in a particular language. Beyond this, input might include information pertaining to the meaning, structural characteristics, or grammatical status of particular utterances; input may be ordered in some manner (e.g. from simple to complex sentences); and so forth. Second, a set of constraints on the modifications that the learner may apply to her grammar must be defined. This set of constraints serves to define the range of possible grammatical hypotheses a learner may entertain and thus, indirectly,

FL014 606

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Stanford Univ.

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

defines a class of possible languages. Finally, a processing mechanism must be defined that determines whether input and grammar are consistent and selects particular modifications when appropriate.

Given characterizations of each of these components, the goal of any particular proof is to determine whether the learner, provided input as defined from any one of the possible languages, must converge on a grammar generating that language. If so, the class of languages under investigation is said to be learnable; otherwise, it is unlearnable. By hypothesis, the class of natural languages is learnable; hence, the construction of learnability proofs concerned with this class typically proceed backwards, by adopting sets of assumptions concerning the three components defined above that will allow positive conclusions to be drawn.

A central difficulty encountered in attempting to construct a feasible learnability proof for natural languages is that the child's language input is strictly limited. Most crucially, the child is not exposed to a representative sample of syntactic constructions from the language to which she is exposed. This is quite clear in the case of sentence embedding. All languages allow sentences with multiple levels of embedding. However, the data from studies by Snow (1972) and Newport (1977), among others, of speech to children indicate that children mostly hear single words, single phrases, or single clauses. Sentences with one level of embedding occur occasionally, but sentences with multiple levels of embedding appear to be very rare indeed, outside of such stylized contexts as "The House that Jack Built." This is a serious limitation, since the nature of rules which operate in simple sentences may be fully revealed only through their interactions with other rules in complex sentences. Thus, one of the empirical constraints on learnability is that, with respect to embedding, input is syntactically simple and, therefore, not maximally informative.

One productive line of inquiry has been concerned with demonstrating how transformational grammars, augmented versions of the type outlined in Chomsky (1965), might in principle be learned, given strict constraints on the complexity of available input. Much of this work is due to Kenneth Wexler and his colleagues. Hamburger & Wexler (1973) outlined the first positive learnability result for transformational grammars; this proof was more fully developed in Hamburger & Wexler (1975). This initial proof conservatively required that the learner be exposed to sentences with up to 400,000 levels of embedding - in the terminology of learnability theory, input of Degree 400,000 - clearly an implausible requirement. In subsequent work, Wexler & Culicover (1980) achieved a much more powerful result: learnability from Degree 2 input. This result entailed the inclusion of a rich set of constraints on grammatical hypotheses. Recently, by adopting a key assumption concerning the information available in input, I have been able to construct a proof demonstrating learnability from Degree 1 input. That is, in this proof, the learner must be exposed to some sentences with one level of embedding; more complex input is not necessary. This is the first learnability result which is fully consistent with empirical data

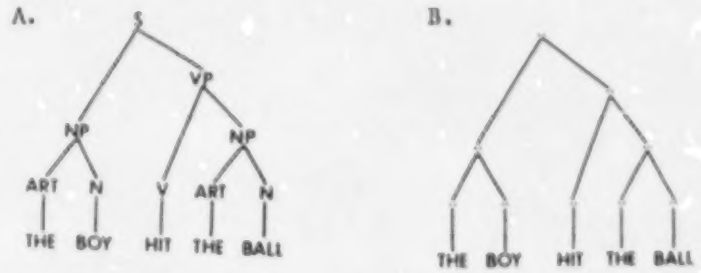
concerning the complexity of children's language input.

The key assumption of the Degree 1 proof is that the learner can reconstruct the surface phrase bracketing of at least some input sentences. This assumption is not that the learner can reconstruct complete surface structures (cf. Figure 2a), but, more simply, that the learner can reconstruct surface bracketing of sentences (cf. Figure 2b). Languages may encode bracketing information in a variety of ways: by prosodic patterns, locations of closed-class tokens, morphological or phonological concord, and so forth, so that this assumption has at least prima facie plausibility. Among the learnability-theoretic consequences of this assumption are that a significantly less rich set of grammatical constraints is required and that learning may proceed in a more efficient manner than was the case in Wexler & Culicover's Degree 2 proof. In addition to demonstrating the theoretical benefits of surface phrase bracketing information, the empirical validity of this assumption must be established; I will return to this issue below. First, I will discuss how the Degree 1 proof avoids the sort of problem which was central to Wexler & Culicover's proof, and I will provide a general outline of this proof. Readers interested in further details should consult Morgan (1983).

Figure 2.

A. Surface Structure

B. Skeletal Structure



Demonstrating learnability from any particular degree of complexity of input is tantamount to showing that any error the learner might make will be apparent in some datum of that degree (Hamburger & Wexler, 1975). The difficulty in Wexler & Culicover's proof is that, given their definition of input, the learner may make errors in simple data that will only become apparent in complex data. That is, it is possible to hypothesize a rule which works perfectly well in simple sentences, but incorrectly in more complex sentences. An abstract example will illustrate how this might occur.

Wexler & Culicover assumed that input consisted of base structure-surface string pairs. One such datum is shown in Figure 3a. In this example, capital letters denote grammatical categories; 'S' stands for 'sentence', but other categories are arbitrary. In order for her grammar to generate the surface string, EGFJH, the learner must hypothesize a transformation which relocates G. One solution is shown in Figure 3b. This is not the only possible solution, but given this datum it is perfectly adequate. Moreover, given only this datum, there is nothing to indicate to the learner that

this solution is incorrect. However, given a more complex datum, in this case a Degree 2 datum, as shown in Figure 3c, it will become apparent that the transformation hypothesized in the simpler datum is incorrect. In generating the proper surface string for this datum, FJHYEG, the previously hypothesized rule must first be applied. The resulting structure is shown in Figure 3d. The problem now is how to get EJ to the end of the string. However, formulating a transformation to do that would entail moving parts of separate constituents, and this is not allowable in any transformational account. In sum, the learner's first transformation ought to have been formulated so that G became a sister of E, rather than of F, but the necessity of this could only become apparent in relatively complex data.

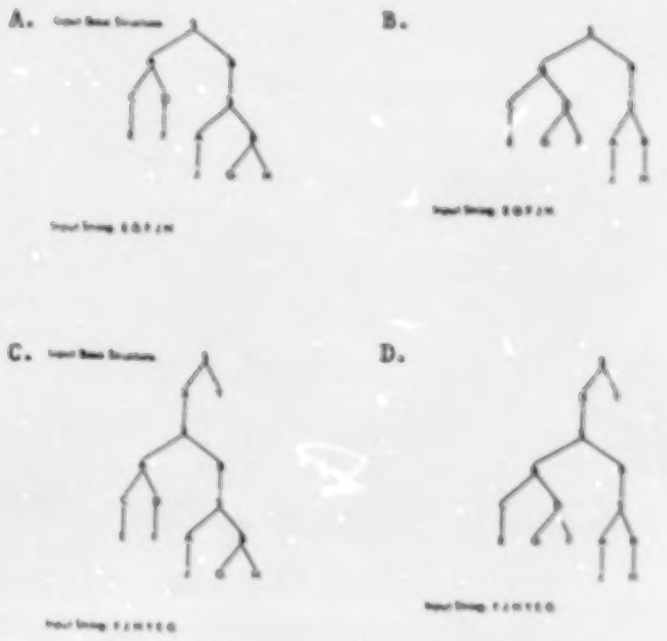
Figure 3.

A. Degree 1 Input Datum.

B. Partially-Derived Degree 1 Datum.

C. Degree 2 Input Datum.

D. Partially-Derived Degree 2 Datum.



The rule interaction illustrated in this example is quite simple. It is possible to have interactions among rules such that errors made on simple data only become detectable in extremely complex data. If learnability is to follow from simple data (in this case, Degree 2), such pernicious interactions must somehow be ruled out. Wexler & Culicover's solution to this problem was to adopt a rather large set of constraints on grammatical hypotheses and rule operation; the bulk of their 1980 book was concerned with providing the learnability-theoretic motivations for and adducing linguistic evidence consistent with these constraints.

An alternative approach is to enrich the assumptions concerning input. Consider the example given in Figure 3a again, supposing that the learner had, in addition, access to information about the phrase bracketing of the input sentence, perhaps provided by prosodic con-

tours or morphological cues. The learner's input would effectively be [EG]FJH (ignoring irrelevant details). As before, the learner must formulate a transformation which moves G. Note that the erroneous solution which was possible in the absence of bracketing information is now impossible, since it would produce the wrong phrase bracketing, and the learner's derived surface structure would be inconsistent with the input datum. Thus, an entire class of errors which was problematic in Wexler & Culicover's work can never occur, given bracketed input. In general, with bracketing information, errors made in simple data will be apparent in simple data. Granting the assumption that rules cannot span intermediate embedded clauses (this is the gist of Wexler & Culicover's Binary Principle), learnability from Degree 1 input would appear to follow directly.

In fact, it does not: bracketing information does not eliminate all types of errors. It is still possible for the learner to formulate rules which apply either undergenerally or overgenerally. Moreover, it is possible to have interactions of rules which result in the learner's erroneous rules applying in the correct range of generality in simple sentences and, say, overgenerally only in complex sentences. As an example, consider the data presented in Figure 4. Suppose that upon receipt of the datum in 4a, the learner formulated the rule X-C-Y-B -> X-C+B-Y. However, this rule would apply overgenerally to the datum in 4b; the correct rule would omit the initial variable. Assuming that the base rules include only those which may be read off these trees, there could be no Degree 1 datum which would reveal this error. Thus, as in Wexler & Culicover's proof, it is necessary to assume a set of constraints on grammatical hypotheses in order to rule out the possibility of such errors and interactions (in this case, the required constraint is that 'S' may only be introduced as an optional category in the base). However, the set of constraints required here is substantially simpler, approximately one-third the size of that required in the Degree 2 proof.

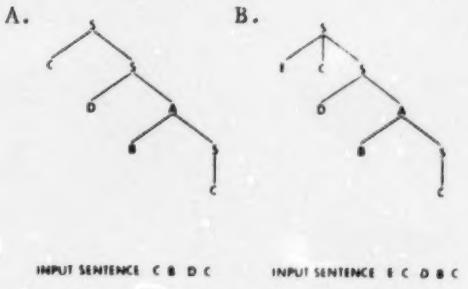


Figure 4.

A. Learner's Rule Applies Correctly.

B. Learner's Rule Applies Overgenerally.

Given this set of constraints on rule operation and formulation, along with input consisting of base structure-bracketed surface string pairs, three essential theorems can be proven. First, the same set of rules applies in sentences with one or no levels of embedding that applies in more complex sentences. Second, any rule which applies over- or undergenerally in complex sentences will apply

in the same manner in Degree 1 sentences. Third, any error which might always be concealed by interactions with other rules in Degree 1 sentences will also always be concealed in more complex sentences (and thus will never be revealed). These three points ensure that any detectable errors the learner might make will be made apparent by some Degree 1 datum. Degree 1 learnability follows from this.

Several learnability-theoretic advantages follow from the assumption that phrase bracketing information in available in language input. First, learning may proceed on the basis of very simple data. More important, the required input complexity is completely consistent with that which empirical studies have shown to be available to children learning language. Second, both because the required data is simple and because the availability of surface bracketing information discourages certain types of errors, learning - that is, convergence on the target grammar - can be substantially more efficient. Clearly, other things being equal, the fewer the errors, the faster learning will progress. In addition, should the learner make an error, the probability of an error-revealing datum appearing in input becomes greater as the required complexity of input declines. This is because the number of possible sentences with a given number of embeddings is a geometric function of that number: there are far fewer Degree 1 sentences than Degree 2 sentences. Third, given bracketing information, many constraints on grammatical hypotheses which would otherwise be necessary may be dispensed with. A number of the constraints adopted in Wexler & Culicover's proof have been criticized (e.g. Williams, 1981) on the grounds that they make incorrect linguistic predictions. Significantly, these constraints are among those which either are not required or can be reformulated to produce more accurate predictions in the Degree 1 proof.

One criticism commonly directed toward Wexler & Culicover (e.g. Pinker, 1981) is that their input assumptions - in particular, their assumption that the learner can reconstruct appropriate base structures for input sentences - are implausibly strong. As should be apparent, the Degree 1 proof is even more vulnerable on this point. However, this criticism overlooks one of the central points of these proofs. It might seem that granting the learner information about both base and surface structures would make the learning of a transformational grammar trivial. But the Degree 1 proof demonstrates that, even given a highly informative language learning environment, acquisition will not uniformly succeed unless there are further constraints on the inductive hypotheses which may be considered.

Still, fewer constraints are required in the Degree 1 proof. This is a rigorous demonstration that the inclusion of certain types of information in input <u>can</u> replace the need for intrinsic constraints on the child's grammatical hypotheses. This point must be interpreted conservatively. While surface bracketing information may reduce the need for constraints on grammatical hypotheses, it will concomitantly increase the need for perceptual and computational capacities required to accurately construct representations of such bracketing. In short, while the availability of additional infor-

mation in language input may change the nature of the contributions of child to acquisition, it is highly doubtful that more information will in any way reduce the overall need for such contributions.

The issue remains whether the assumption of phrase bracketing information is empirically justified. Two basic questions relevant here are whether there are sufficient cues to phrase bracketing in children's language environments and whether young children have the processing abilities to accurately perceive and interpret such cues. I will address these questions briefly below.

As noted above, languages potentially include many cues to phrase bracketing, including prosodic patterns, phrase-peripheral placement of function words, and phenomena of concord. Additional cues may be available to young children: they often hear examples of phrase in isolation or sequences of partial-plus repetitions in which phrases may be systematically permuted or substituted (Newport, 1977). One phenomenon commonly noted in studies of speech to children is that the prosody of such speech is "exaggerated." It is possible that cues to phrase boundaries (e.g. final vowel lengthening, pitch discontinuities at phrase junctures; cf. Cooper and Paccia-Cooper, 1980) may be among those aspects of prosody which are reliably exaggerated. In an initial study investigating this possibility, Jan Goodsitt and I constructed a number of sentence pairs in which a CVC target occurred in either phrase-final or non-position, e.g., "He climbed to the top and he stared" and "He climbed to the top of the stairs." We embedded these sentences in a children's story and had mothers read this story both to an adult experimenter and to their 2- to 4-year-old children. Preliminary acoustic analyses of the data indicate that the duration of phrase-final targets is reliably longer in speech both to adults and to children; however, the difference between phrase-final and non-phrase-final targets is both absolutely and proportionately greater in speech to children. This is a preliminary indication that prosodic exaggerations may provide children with clues to language structure. However, whether these cues, prosodic and otherwise, encode sufficiently detailed information about the hierarchical structure of phrases is an open question.

There is little direct evidence concerning the extent to which children are sensitive to the structural connotations of the cues noted above. In a study involving 3- to 5-year-olds, Poldi Gerard-Ngo, Jan Goodsitt, and I induced children to play a game involving the echoing of final portions of nonsense sentences. Children were first exposed to a model that echoed words that either fall within a prosodically-defined phrase or span two such phrases. The number of words was the same in either case. The results are quite clear-cut: regardless of the type of model to which children were exposed, they had a strong tendency to echo a complete prosodic phrase. Whether the short term memory units in evidence in this study are equivalent to syntactic units is an issue requiring further investigation.

A third question relevant to the evaluation of the bracketed input hypothesis is whether availability of phrase bracketing infor-

mation influences the child's selection of grammatical hypotheses. This issue is elusive: since cues to phrase bracketing are apparently universally present in language input, it is not possible to correlate their presence or absence with particular acquisition outcomes. One possibility is to investigate acquisition of bracketing analyses of constructions for which distributional evidence is lacking. Different learners would be expected to come to a consistent analysis only if input bracketing information were guiding their hypotheses. For an initial study along these lines, see Morgan (1983). However, the universality of cues to phrase bracketing is in itself suggestive that such cues may play an important role in acquisition.

In sum, the Degree 1 learnability result described here indicates that in the acquisition of complex grammars, such as those of natural languages, the required complexity of input can be reduced and constraints on possible grammatical hypotheses can simultaneously be relaxed, but only if the input available to the language learner is rich in structural information.

REFERENCES

Chomsky, N. Aspects of the theory of syntax. Cambridge, MA: MIT Press, 1965.

Cooper, W. E. & Paccia-Cooper, J. Syntax and speech. Cambridge, MA: Harvard University Press, 1980.

Gold, E. M. Language identification in the limit. Information and control, 1967, 10, 447-474.

Hamburger, H. & Wexler, K. Identifiability of a class of transformational grammars. In K. J. J. Hintikka, J. M. E. Moravcsik, & P. Suppes (Eds.), Approaches to natural language. Boston: D. Reidel, 1973.

Hamburger, H. & Wexler, K. A mathematical theory of learning transformational grammar. Journal of Mathematical Psychology, 1975, 12, 137-177.

Morgan, J. L. Learning a complex language from simple input: Contributions of phrase bracketing information. Unpublished Ph.D. dissertation, University of Illinois, 1983.

Newport, E. L. Motherese: The speech of mothers to young children. In N. Castellan, D. Pisoni, & G. Potts (Eds.), Cognitive theory: vol. II. Hillsdale, NJ: Erlbaum Associates, 1977.

Pinker, S. What is a language, that a child may learn it, and a child, that he may learn a language? Journal of Mathematical Psychology, 1981, 23, 90-97.

Snow, C. E. Mothers' speech to children learning language. Child Development, 1972, 43,549-565.

Wexler, K. & Culicover, P. W. Formal principles of language acquisition. Cambridge, MA: MIT Press, 1980.

Williams, E. S. A readjustment in the learnability assumptions. In C. L. Baker & J. J. McCarthy (Eds.), The logical problem of language acquisition. Cambridge, MA: MIT Press, 1981.